DOCUMENT RESUME

ED 026 454 VT 001 818

By-Gilbert, Ardyce Lucile

Clinical Evaluation of Predictive Data for Prospective Home Economics Teachers.

Iowa State Univ. of Science and Technology, Ames.
Pub Date 66
Note-44p.
EDRS Price MF-$0.25 HC-$2.30
Descriptors-Educational Planning, Effective Teaching, Evaluation Criteria, *Home Economics Education, *Home Economics Teachers, Longitudinal Studies, *Predictive Measurement, Rating Scales, *Success Factors, Teacher Characteristics, Teacher Education, Teacher Evaluation

This investigation, part of a longitudinal study of homemaking teacher effectiveness, was designed to explore the usefulness of clinical judgments to predict teacher success. Clinical judgment is defined as involving the ability to make sound decisions after gathering and evaluating all the pertinent evidence, weighing possible alternatives in terms of past experience or normative probabilities, and arriving at problem solutions which reflect basic science orientations. The plan worked to determine the reliability of the judge's estimates and to correlate their estimates and the composite success scores. Ten judges, including clinical psychologists, guidance counselors, and home economic teacher educators each analyzed 16 randomly assigned cases, providing two evaluations per case. Statistical analysis revealed significant differences among judges, subjects, and measures, and that correlation of judges' estimates and composite success scores was not feasible. All types of predictive data were considered useful as they were referred to in the judges' evaluation. The appendix contains interpretations of test scores and samples of the judges' rating sheet. (FP)

ED026454

4a(15

CLINICAL EVALUATION OF PREDICTIVE DATA FOR PROSPECTIVE HOME ECONOMICS TEACHERS

Ardyce Lucile Gilbert

VTC01818

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Clinical Evaluation of Predictive Data For Prospective Home Economics Teachers

by

Ardyce Lucile Gilbert

Directed by

Hester Chadderdon, Ph.D.
Professor, Home Economics Education

Thesis conducted as part of Project No. 15, Prediction of Success of Graduates of Iowa State University in Teaching Vocational Home Economics, under grant from Iowa Department of Public Instruction, Division of Vocational Education, PL88-210 Sec. 4(a).

Iowa State University
of Science and Technology
Ames, Iowa
1966

CLINICAL EVALUATION OF PREDICTIVE DATA FOR PROSPECTIVE HOME ECONOMICS TEACHERS

by

Ardyce Lucile Gilbert

A Thesis Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
MASTER OF SCIENCE

Major Subject: Home Economics Education

Approved:

In Charge of Major Work

Head of Major Department

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1966

TABLE OF CONTENTS

	Page
INTRODUCTION	1
REVIEW OF LITERATURE	3
METHOD OF PROCEDURE	11
FINDINGS AND DISCUSSION	21
SUMMARY	27
LITERATURE CITED	30
ACKNOWLEDGEMENTS	32
APPENDIX A. INTERPRETATIONS OF TEST SCORES	33
APPENDIX B. JUDGE'S RATING SHEET	37

T18574

LIST OF TABLES

		Page
Table 1.	Graduates of Iowa State University who taught Homemaking I and/or II classes from 1961-1962 to 1965-1966	12
Table 2.	Analysis of variance of judges, subjects, and measures	22
Table 3.	Number of subjects placed at each degree on the certainty scale by judges	23
Table 4.	The number of point differences between first estimates of judges	24
Table 5.	Number of times estimates of judges disagreed by three or more points	25

INTRODUCTION

For many years researchers and educators have been aware of the responsibility for and the importance of the decisions involved in accepting candidates for a teacher education program. Studies have been conducted with the hope of establishing a sound basis for judgments but those designed to predict the effectiveness of teachers have produced disappointing results. One possible reason for failure is that the statistical analysis of the predictive data is inadequate. In a few studies involving prediction clinical judgments were used to advantage; hence, it seemed feasible to try this method of analysis in predicting teacher success. Some clinical psychologists believe that the clinical is superior to the statistical in predicting behavior because of the human element involved.

A longitudinal research project[1] is currently being conducted at Iowa State University to predict effectiveness of homemaking teachers. The selection of criteria and predictive data was begun in 1958 and, in a recent exploratory study, Crabtree (3) used a statistical analysis to investigate the relationship of the prediction and the criterion measures. The correlations between the two were positive but inadequate

---

[1]Iowa State University Agricultural and Home Economics Experiment Station Project 1413

for prediction of teacher success on an individual basis. The present study was designed to explore the usefulness of clinical judgments to predict teacher success.

Several interpretations of the term "clinical judgments" are found in the literature but Thorne's definition has been accepted as pertinent to the present study:

> Clinical judgment is operationally defined as involving the ability to make sound decisions after gathering and evaluating all the pertinent evidence, weighing possible alternatives in terms of past experience or normative probabilities, and arriving at problem solutions which reflect basic science orientations (the cultural value system against which scientists operate). (14, p. 128)

REVIEW OF LITERATURE

The concern of educators for many years with the quality of teachers being educated has led to numerous investigations in an effort to identify and measure characteristics of a successful teacher. Several have been directed toward prediction of teacher effectiveness but none was found that used the clinical method of analyzing data collected for prediction. The studies here will be limited to those concerned with validity and reliability of clinical judgments and clinical predictions of performance based largely on non-projective types of tests. Projective data differ so greatly in nature from those used in the present investigation that research based largely on them was eliminated. Investigations involving the prediction of teacher effectiveness at the secondary level will include only recent investigations to supplement Crabtree's review.

The use of clinical judgments in predicting performance has been a controversial issue for some years, particularly among clinical psychologists and research workers in that field. Meehl (10) was among the first to question clinical judgments as being too subjective. Although more recently (11) he appears to have modified his view, he continues to support the actuarial method as the only sound one. The assertion of several writers, including Meehl, that few clinicians have demonstrated validity and that they make no

better than chance predictions has caused other clinical psychologists to retaliate. Holt (6) and Thorne (14) base their objections to these arguments on the premise that the studies cited compared the performance of the best tests with the predictions of unspecified groups of clinicians. Thorne points out that

> research which samples the averaged judgment of good and bad clinicians tends to produce no better than chance prediction because the superior judgments of the good clinicians are balanced off by the invalid judgments of the poor. The crucial test of Meehl's hypothesis is to compare the judgments of the best clinicians with the best actuarial predictions. (15, p. 116)

Holtzman (7) suggested that the relevance of either method can be determined only by analysis of the activities involved, and that dissention arises because the supporters of either method tend to oversimplify the problem. He implied that both methods are valuable in their own context, the actuarial for data processing and the clinical for interpretations where the human element cannot be eliminated.

In an attempt to substantiate his belief that clinical judgments are valid, Newton (12) divided 50 subjects into five equal groups consisting of the socially adequate, non-hospitalized neurotics, and hospitalized schizophrenics. A seven-point quantitative scale of adjustment was constructed to aid in rating the subjects. Each of 10 psychiatrists and 10 psychologists was asked to evaluate the clinical materials of 15 subjects and to rate them in terms of over-all adjustment.

Newton found a high degree of reliability among the psychiatrists judgments (.91) and the psychologists judgments (.94), and a significant relationship (.86) between the judgments of both disciplines.

Some of the criticisms regarding validity of clinical judgments were directed at the unfamiliarity by the clinicians with the criteria involved and the use of rating scales with which they had had little experience. Lewinsohn et al. (9) investigated the validity of clinical interpretations from a battery of psychological tests commonly used in practice. Five psychologists gave blind ratings of test protocols for a randomly selected sample of 100 psychotic and neurotic hospital patients. The ratings were based on a battery of tests, age, and sex of patient and recorded on a 23-item rating scale which the psychologists had previously helped to develop and with which they had had experience. Each judge rated 40 subjects, thus providing two independent judgments for each patient. The criteria were parallel ratings based on the patient's hospital chart and on an interview by a psychologist who was unaware of the test results. The authors report that "validity coefficients obtained were predominately in the direction of supporting the validity of the test ratings" but that "the validity differed with different areas of patient functioning."

Instead of using interview data as the criterion measure, Bobbitt and Newman (2) employed ratings based, in part, on

interviews as one of three bases for predicting the success of officer candidates in the United States Coast Guard Academy. Each officer candidate was interviewed by a psychologist and a psychiatrist from the medical department of the Academy, who were provided with the Personal Data Questionnaire and available test scores for the candidate. The interviewer assigned an overall rating based on a written summary of his evaluation and interpretation of the test scores for each candidate. A second prediction measure was the combined scores for three of the tests which had been available to the interviewer: 1) quantitative ability, 2) verbal ability, and 3) bi-dimensional spatial perception. The third basis for prediction was the combined scores of the interviewers rating and the tests just enumerated. The criterion to be predicted was the degree of success during training in the Reserve Training School, based on academic achievement and adaptability records which were considered in the final class standing. These were used to classify the candidates into four groups. Bobbitt and Newman reported that there was a direct relationship between the prediction measures and the success of the officer candidates, but that the combination of the interview ratings and the test scores produced a better prediction than either did separately.

In a prediction of performance of aviation cadets, Holtzman and Sells (8) investigated the possibilities of developing a hypothesis for the quantitative scoring of a group of screening tests. A group of 100 cadets, 50 who had made

successful adjustments to flight training and 50 who had been unsuccessful in completing the program, were randomly selected and an experimental design was developed wherein each of 19 psychologists rated 20 cadets by two methods: judgment based on one test at a time, and judgments based on a global evaluation of the predictive data available for each subject. In addition, each judge was asked to state the cues which influenced his evaluation and to indicate on a three-point scale the degrees of confidence he had in the validity of his judgments. Little relationship between the clinical evaluation of the cadets and the measure of flying success was found. However, the amount of agreement among the judges for the global approach tended to be significantly better than chance.

In three recent studies attempts have been made to predict teacher competence. In search of college records which might be used to predict success Freehill (4) investigated the relationship of college recommendations and field evaluations. The college recommendations were based on the records which included: 1) entrance test data, 2) academic records, and 3) a report on social and community life which was rated on a 10-point scale indicating the student's degree of strength or weakness as a teacher candidate. The entrance test data included scores from an academic aptitude examination, three English test scores, and 10 scores from instruments developed in the American Council on Education Study of Evaluation. The

academic record was the grade point average computed for eight subject-matter groupings, and the rank relative to social and community life was based largely on ratings made by faculty and supervisors of student teaching. The field evaluations were obtained near the end of the first year of teaching and again near the end of the fifth. Principals, superintendents, and supervisors or vice principals rated the teacher on the basis of professional and personal qualities at these two periods. Freehill reported that there was a positive relationship between the principal's judgment and the college recommendations, but that on professional qualities the later field evaluation agreed more with the college evaluation than did the earlier field evaluation.

Sprinthall, et al. (13) ignored the "static personal traits which cannot be made operational" and concentrated on a conceptual framework for prediction and evaluation of teaching success based on observable teacher classroom behavior and its relationship to cognitive flexibility-rigidity scores of psychological tests. Twenty-eight subjects were randomly selected from a population of graduate students enrolled in a master of arts teaching program which involved one year of study. During the summer, seven weeks were devoted to intensive supervised student teaching, while half of the following academic year was concentrated on full-time classwork and the other half on intern supervised teaching in a local school system. The predictive data were obtained from two

psychological tests, Rorschach and Visual Impression Test (VIT), which were administered before the subjects began student teaching. The criterion measure consisted of ratings of teacher behavior based on observations during a 60-minute sample of the student teaching period and subsequent supervisory-planning conferences. The results indicated that effective teaching and cognitive flexibility-rigidity are related. Sprinthall et al., however, suggested that the predictive data need to be refined and that complete follow-up information of success as a full-time teacher be made to validate the criterion measure.

In a study of homemaking teachers Crabtree (3) analyzed data that had been collected to determine the relationships between selected predictors and success criteria using subjects who were graduates of Iowa State University. During their enrollment in the University predictive data were collected by a battery of instruments: The Guilford-Zimmerman Temperament Survey (GZTS) and the Minnesota Counseling Inventory (MCI) to measure personality traits; the Just Suppose Inventory (JSI) to indicate certain attitudes; and the Johnson Home Economics Interest Inventory (JHEII) to obtain an estimate of vocational interests. Also included was the cumulative quality point average (CQPA) at the end of the sophomore year. The three criteria used to determine the effectiveness of a teacher were teacher-pupil rapport, pupil gain in ability to apply generalizations in solving problems in home economics, and the

adjustment of the teacher to school and community. Sixty-four homemaking teachers were included in her analysis. A panel of judges rated the predictors and criteria in terms of their relative importance and an adaptation of the J-coefficient procedure was used to provide weights for each. The weighted predictors were summed to obtain composite prediction scores and, similarly, composite criteria scores were secured. When these were intercorrelated she found that academic achievement had a significant but low correlation with the composite criterion score. Also scores on attitudes toward low-income groups and toward middle and upper class groups correlated positively with the composite criterion and all individual criterion scores. Although several scores from the predictive data exhibited positive correlations with the composite success scores, the composite scores were not high enough to use in the prediction of teaching success for an individual.

METHOD OF PROCEDURE

Purpose of Study

This study is part of a longitudinal study to predict the effectiveness of homemaking teachers who are graduates of Iowa State University. Since the statistical approach used by Crabtree did not produce satisfactory estimates for the prediction of the success of an individual student, a clinical analysis was explored using the predictive data available for 80 first-year teachers. The present study was designed also to determine the number and the type of judges needed to make reliable estimates.

Description of Population

An attempt was made to obtain data for all graduates who taught in Iowa during the period 1961-1962 to 1965-1966. Because the achievement tests used to measure pupil gain were based upon the Iowa Homemaking Curriculum Guides, only those graduates who taught in Iowa schools were included. Since most first-year graduates teach classes at the ninth- or tenth-grade level, it was not feasible to develop instruments to measure success of the few who taught only at other levels. For this reason the population is further limited to those teaching Homemaking I and II classes in Iowa.

Table 1 presents information concerning the number of home economics education graduates of Iowa State University

from 1961 to 1966 who taught Homemaking I and/or II classes in Iowa, and the reasons for excluding a portion of the population from this study.

Table 1 Graduates of Iowa State University who taught Homemaking I and/or II classes from 1961-1962 to 1965-1966

Graduates	1961-'62	1962-'63	1963-'64	1964-'65	1965-'66	TOTAL
Included in study	3	17	15	29	16	80
Incomplete predictive data	17					17
Refusal to cooperate*	2		3	1	2	8
Resignation before end of year	1				1	2
Incomplete success data**		6	6			12
Errors in administering test			1			1
Late placement of teacher		1			1	2
TOTAL	23	24	25	30	20	122

*Superintendent or teacher
**Largely incomplete administrator's ratings

Incomplete data were available for a large number of the 1961-62 graduates because the second personality inventory was selected for use too late to be administered to most of these students.

Of the 80 teachers for whom complete data were available 54 taught both Homemaking I and II classes, 21 taught Homemaking I but not Homemaking II classes, and 5 taught only Homemaking II classes.

Predictive Data

These data included the cumulative quality point average and a battery of four instruments. The Guilford-Zimmerman Temperment Survey (GZTS) and the Minnesota Counseling Inventory (MCI) were employed to obtain an estimate of personality traits, and the Johnson Home Economics Interest Inventory (JHEII) to determine vocational interests. The Just Suppose Inventory (JSI), which is not yet published, involves attitudes toward other persons and groups.[2] The student is asked to project herself into each of the 12 situations which might be encountered by a teacher and to select statements which reveal her attitudes. The situations relate to: acceptance of changing conditions in our society, especially broken homes and mothers working and of parents with little or much education; adaptability to communities of different sizes and to various areas within a community, i.e. industrial sections of a city, slum districts, suburban areas; tolerance of

---

[2]Copies are on file in Department of Home Economics Education, Iowa State University. Permission was obtained from Ruth Lehman, Ohio State University, to use this inventory.

foreign-born and ethnic groups other than one's own; respect for different religions and for families in the low, middle, and upper income groups; understanding problems involved when working with low I.Q. or deliquent students, persons living in a three-generation family home; and attitudes toward parents in relation to concern about their children's welfare.

The students who entered Iowa State University as freshmen were administered two of the inventories either at the end of the freshmen year or the beginning of the sophomore year and the other two at or near the end of their sophomore year. Students who transferred into the University or into the Home Economics Education Department reacted to these inventories soon after the transfer. For the students entering as freshmen the cumulative quality point average was recorded when they were formally admitted to the home economics teacher program, commonly at the end of their sophomore year. For the transfer students this average was recorded when they applied which usually was after the completion of two quarters of work at Iowa State University.

These data were supplemented in the present study with information concerning pre-college work experiences and activities, a statement by the advisor of the student's strengths and limitations, and the student's statement of motivation to teach. These were obtained from the Application for Admission to Teacher Education Curriculum in Home Economics. Also included were statements obtained from the

"Instructors Report" which indicate characteristics of behavior observed by the classroom teachers in the College of Home Economics. It was hoped that the work experiences and the activities would give some indication of leadership qualities, that the statements of the advisors and the teachers would supplement information concerning the ability to relate to other persons and that the student's statement would provide further insights into the personality and abilities of the prospective teacher.

Prediction of Success

Ten judges analyzed the predictive data and made estimates for the 80 subjects. Each judge was provided the information previously described and also some data to aid in interpretations.

For the GZTS a summary[3] was developed from the Manual of Instructions and Interpretations (5) of the qualities which describe a high and a low scorer for each of the 10 personality characteristics. In addition the scores of each subject were recorded on the profile based on the responses of 389 college women reported in the manual. These data were supplemented by means and standard deviations derived from a sample of 100 sophomores in the Department of Home Economics

---

[3]A copy of Interpretation of Scores - Extremes may be found in Appendix A.

Education. A clear plastic overlay containing the mean and the ± one standard deviations from the mean was supplied each judge.

Descriptions of the high and the low scorers for the personality traits measured by the MCI were duplicated from the manual (1). The scores of the subjects were recorded on a profile based on the response of 367 sophomore, junior, and senior students who were majors in home economics education.

The profile for the JHEII has been published[4] but since it is based on scores of freshmen home economics students, a clear plastic overlay indicating the means and the ± one standard deviation from the means based on a sample of 100 sophomores enrolled in home economics education was provided to facilitate analysis of the scores on this Inventory.

No manual is available for the JSI, therefore an explanation of the attitudes which might be expressed by high and low scorers was made from the statements included in the Inventory.[5] A profile which had previously been developed, based on the response of 330 sophomores enrolled in home economics education, was used to record the score of each subject.

Of the 10 judges selected, two were clinical psychologists, five were guidance counselors, and three were staff

---

[4]Published by The Iowa State College Press, Press Building, Ames, Iowa.

[5]A copy of the Just Suppose Inventory: Interpretation of Scores may be found in Appendix A.

members of the Home Economics Education Department. Each judge analyzed 16 cases which had been randomly assigned; this provided two estimates for each case. An eleven-point scale was used to determine the degree of certainty of the estimations.[6] The judges were directed to base their estimates on the likelihood of the student being successful in a one-teacher department in a high school with an enrollment of less than 400 students, located in a relatively small Iowa town, a population of 1200-7000, with few lower-class families. This description was based on information concerning the most frequent teaching situations of the teachers involved in the study. It was assumed that the majority of home economics education graduates of Iowa State University would be employed in similar situations. Because a few of the graduates taught in urban communities, a second estimate was also made relative to success in a larger urban school. In addition, the judges were asked to explain any score below 6, with the hope that the explanation would be useful in understanding differences among judges and in determining which data to continue to collect.

Success Data

The three criteria used in the project for determining

---

[6]A copy of the Judge's Rating Sheet may be found in Appendix B.

teacher effectiveness were: pupil gain in ability to apply generalizations, pupil-teacher rapport, and teacher adjustment to school and community.

Pupil gain was determined by two forms of two achievement tests. Form A, administered at the beginning of the school year, and Form B, administered near the end of the school year to Homemaking I and II classes. The tests were developed by the project leader to assess the gain in ability to apply generalizations in solving problems in home economics. The Homemaking I test included five areas of homemaking: food and nutrition, textiles and clothing, child development, family relations, and housing. The Homemaking II tests included these with the exception of child development since the state curriculum guide did not contain a unit in this area at this level. A class mean for Homemaking I classes was computed for each teacher by subtracting the sum of the scores on Form A from the sums on Form B, and dividing by the number of pupils who completed both forms; the same procedure was used to obtain a mean for Homemaking II classes.

Pupil-teacher rapport data were collected by administering two forms of the SETC inventory, one for Homemaking I and one for Homemaking II classes. These inventories consist of statements about the homemaking teacher and class to which the pupil indicates his feelings by agreeing or disagreeing. The items relate to the teacher's interest in, understanding of, and attitudes toward the pupil, her willingness to help,

and the amount or kind of help given to the pupil. Each favorable response was given a value of one; all other responses a value of zero. Class means for each teacher were computed by summing the scores for the classes in Homemaking I and in Homemaking II and dividing each by the total number of pupils.

An estimate of teacher adjustment to school and community was made by the school administrator and recorded on a special form designed for this purpose. The factor analysis made by Crabtree yielded two single-item factors and two clusters of items. The former items were Physical Health of Teacher and Judgment in Discussion of Personal and Professional Problems. The two clusters involved Management of Department and Relations with School Personnel, Pupils, and Community. Numerical values of 1-6 were assigned to the responses and numerical values of the clusters were obtained by summing the scores for the responses to the items contained in the cluster.

Treatment of Data

In the present investigation the estimates of 10 judges were tabulated and the data were analyzed to obtain the degree of variance due to judges, subjects and measures.

A reliability coefficient was computed for one judge using the following formula:

$$R_j = 1 - \frac{\sigma_e^2}{\sigma_e^2 + \frac{\sigma_c^2 - \sigma_e^2}{nN}}$$

Where

R_j = reliability of one judge

σ_e^2 = error variance

σ_c^2 = subject variance

n = number of judges

N = number of estimates for each judge

An estimate of the reliability coefficient for ten judges employed the formula:

$$\overline{R}_j = \frac{nR_J}{1 + (n - 1) R_j}$$

$\overline{R}_j$ = reliability of 10 judges

R_j = reliability of one judge

n = number of judges

FINDINGS AND DISCUSSION

To investigate the usefulness of clinical judgments to predict teacher effectiveness, two analyses of data were planned: 1) to determine the reliability of judges' estimates and 2) to correlate their estimates and the composite success scores.

An analysis was made to determine wherein there was variance due to judge, subject, measure, and the interaction between them. The measures are the two estimates each judge assigned the subject as he evaluated her effectiveness as a homemaking teacher. The results are presented in Table 2. The F values were calculated using the error mean square as the denominator, with the exception of M, for which the variance of the interaction of CM was used. The analysis yielded F values that are highly significant for judges, subjects, and measures.

Since one of the purposes of the present study was to explore the accuracy of clinical judgments, reliabilities of judges were computed. A reliability coefficient of .142 was obtained, thus theoretically limiting the validity coefficient of the judgments to .37. Because the estimates of ten judges were employed in the present investigation, a reliability coefficient for the ten was estimated. The result was a reliability coefficient of .623, which could not yield a validity coefficient above .79. This finding would indicate that even

Table 2 Analysis of variance of judges, subjects, and measures

Source of variation	Sum of squares	d.f.	Mean square	F value
Subjects (C)	720.620	79	9.122	4.30*
Measures (M)	20.503	1	20.503	23.06*
C M	70.246	79	.889	.42
Judges (J)	165.653	9	18.406	8.68*
J M	6.903	9	.767	.36
Error	300.917	142	2.119	
Total	1284.842			

*Significant beyond .001 level

though the ten judges evaluated each case and there was no error variance in the criterion measures, the validity of prediction of success for an individual would be inadequate. In order to increase the reliability and the validity coefficients of the judges' estimates, at least twice as many cases and/or judges need to be included. This, however, is not practical or feasible because of the time involved for evaluating each case. In view of the evidence revealed in the analysis of variance and the reliability of the judgments it was decided not to complete an analysis of the judges' estimates and the composite success scores of the subjects.

Since the reliability of the judges was low, the data were examined to discover possible explanations. The judges

had been directed to give two estimates of success for each subject, one for teaching in a small community and the other in a larger urban area. Both estimates were designated by a numerical rating on the eleven-point certainty scale. The distribution of the judges' first estimate is presented in Table 3. It reveals a skewed distribution; the judges tended

Table 3 Number of subjects placed at each degree on the certainty scale by judges

Judges	Degrees of certainty										
	0	1	2	3	4	5	6	7	8	9	10
A					3	2	1	5	4	1	
B			2			2	2	2	3	3	2
C			2		1	2	3	2	4	2	
D				2	4	3	6	1			
E				1	2	1	3	2	6	1	
F			1	1	1			2	4	2	5
G					2	3	2	1	6	1	1
H							3	4	5	3	1
I						2	4	7	3		
J						1		6	6	3	
Total	0	0	5	4	13	16	18	37	42	16	9

to rate more subjects toward the upper end of the continuum. This was to be expected, however, since the sample included only those students who had been screened before admission to

the teacher education program. The estimates of three judges were markedly different. Judges H and J tended more than the others to use the upper end of the continuum and Judge D the middle.

An analysis of the first estimates of the pairs of judges for each case revealed that all judges differed at least three points on one or more cases. One possible explanation for these differences might be the lack of the judges' experiences with the certainty scale.

Table 4 The number of point differences between first estimates of judges

Judges	Point differences between judges first estimates									
	0	1	2	3	4	5	6	7	8	9	10
A	5	4	4	3							
B	5	6	3			2					
C	4	5	2				2				
D	4	3	1	4	2			2			
E	3	4	4	2	2		1				
F	2	3	5	2	1		1	2			
G	3	4	8	1							
H	6	4	3	1	1		1				
I	6	5	4		1						
J	4	4	1	3	3	1					

The data in Table 5 indicate that Judges D, F, and J exhibited this point difference more than any other judge on both the first and second estimate. It is suggested that the cases estimated by judges D, F, and J be re-evaluated by the judges who less frequently disagreed and another analysis be made to determine reliability.

When a judge assigned an estimate of five or less, he was asked to indicate the reasons for the decision. Upon examining these, it was found the judges who agreed on the estimates for a case tended to select similar bases. The reasons

Table 5 Number of times estimates of judges disagreed by three or more points

Judge	Estimate 1	Estimate 2
A	3	2
B	2	3
C	2	4
D	8	6
E	5	1
F	6	7
G	1	0
H	3	3
I	1	4
J	7	7

given by the judges made reference to all of the types of data given them; hence, it appears that all are useful in prediction. In one case a judge suggested the need for additional information, i.e. age of subject and size of community in which she had lived.

SUMMARY

This investigation is part of a longitudinal research project being conducted to predict the effectiveness of homemaking teachers who are graduates of Iowa State University. The purpose of the present study is to explore the usefulness of clinical evaluations for prediction since a recent statistical analysis of the data revealed the prediction formula inadequate for reliable estimates of an individual. The plan was to determine the reliability of the judges' estimates and to correlate their estimates and the composite success scores.

Predictive data collected in the longitudinal study, which were available for 80 first-year homemaking teachers, included the cumulative quality point average and a battery of four instruments: the Guilford-Zimmerman Temperament Survey and the Minnesota Counseling Inventory to measure personality traits; the Johnson Home Economics Interest Inventory to indicate vocational interests, and the Just Suppose Inventory to determine attitudes toward other persons and groups. These data were supplemented with information concerning pre-college work experience and activities, an estimate by the advisor of the student's strengths and limitations, and the student's statement of motivation to teach.

Each of ten judges, including clinical psychologists, guidance counselors, and members of the Home Economics Education Department, analyzed 16 randomly assigned cases, thus

providing two evaluations for each case. An eleven-point scale was used to determine the degree of certainty of the estimation. The judges were asked to evaluate the student twice, as a teacher in a small community and also in a larger urban area. In addition they were to indicate reasons for a score less than 6, with the hope that the explanation would be useful in understanding differences among judges and in determining which data to continue to collect.

An analysis of variance yielded statistically significant differences among judges, subjects, and measures beyond the .001 level. The reliability coefficients computed for one judge, .142, and estimated for ten judges, .623, indicate that a correlation of the judges' estimates and the composite success scores, as previously planned, was not feasible.

Further examination of the data revealed that the judges' estimates tended to be placed near the upper end of the certainty scale, which was not surprising due to the screening process for admission to the teacher-education program. Lack of experience with the use of the certainty scale may have influenced the judges to make estimates that differed three or more points on a case, however, three judges exhibited this point difference more than other judges. It was suggested that the cases estimated by them be re-evaluated by the judges who less frequently disagreed and another analysis be made to determine reliability.

It appears that all of the predictive data are useful

since reference was made to all types in the reasons given by the judges.

LITERATURE CITED

1. Berdie, R. F. and Layton, W. L. Minnesota Counseling Inventory Manual. New York, New York, The Psychological Corporation. 1957.

2. Bobbitt, J. M. and Newman, S. H. Psychological activities at the United States Coast Guard Academy. Psychological Bulletin 41:568-579. 1963.

3. Crabtree, Beverly. Predicting and determining effectiveness of homemaking teachers. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1965.

4. Freehill, M. F. The prediction of teacher competence. Journal of Experimental Education 31:307-311. 1963.

5. Guilford, J. P. and Zimmerman, Wayne S. The Guilford-Zimmerman Temperament Survey, Manual of instructions and interpretations. Beverly Hills, California, Sheridan Supply Company. 1949.

6. Holt, R. R. Clinical and statistical prediction, A reformulation and some data. Journal of Abnormal Social Psychology 56:1-12. 1958.

7. Holtzman, W. H. Can the computer supplant the clinician? Journal of Clinical Psychology 16:119-122. 1960.

8. Holtzman, W. H. and Sells, S. B. Prediction of flying success by clinical analysis of test protocols. Journal of Abnormal and Social Psychology 49:485-490. 1954.

9. Lewinsohn, P. M., Nichols, R. C., Pulos, L., Lomont, J. F., Nickel, H. J., and Siskind, G. The reliability and validity of quantified judgments from psychological tests. Journal of Clinical Psychology 19:64-73. 1963.

10. Meehl, P. E. Clinical vs. statistical prediction. Minneapolis, Minnesota, University of Minnesota Press. 1954.

11. Meehl, P. E. The cognitive activity of the clinician. American Psychologist 15:19-27. 1960.

12. Newton, R. L. The clinician as judge: total Rorschach and clinical case material. Journal of Consulting Psychology 18:248-250. 1954.

13. Sprinthall, N. A., Whitely, J. M., and Mosher, R. L. A study of teacher effectiveness. Journal of Teacher Education 17:93-106. 1966.

14. Thorne, F. C. Clinical judgment: a clinician viewpoint. Journal of Clinical Psychology 16:128-134. 1960.

15. Thorne, F. C. Editorial comment. Journal of Clinical Psychology 16:115. 1960.

ACKNOWLEDGEMENTS

The writer wishes to express sincere appreciation to Dr. Hestor Chadderdon for her guidance, encouragement, and patience throughout the entire study.

Appreciation is also expressed to the Agriculture and Home Economics Experiment Station for the opportunity to serve as a graduate assistant and to participate in the research of Project 1413.

Gratitude is extended to the following persons, whose help facilitated the completion of the study: Dr. Leroy Wolins for his assistance in the analysis of data; Dr. Russell Canute for his help in obtaining judges for the study; all the faculty who served as judges of data; and the administrators, teachers, and pupils who cooperated in the collection of data.

APPENDIX A. INTERPRETATIONS OF TEST SCORES

GUILFORD ZIMMERMAN TEMPERAMENT SURVEY

INTERPRETATION OF SCORES - EXTREMES

G - GENERAL ACTIVITY
- High Score - strong drive, energy, activity, vitality, speed, courage, enthusiasm
- Low Score - deliberate, inefficient, inactive, slow

R - RESTRAINT
- High Score - deliberate, consistent, self-control restraint, seriousness
- Low Score - impulsiveness, happy-go-lucky, loves excitement

A - ASCENDANCE
- High Score - social boldness, self-defense, leader, being conspicuous, bluffing
- Low Score - social submissiveness, follower

S - SOCIABILITY
- High Score - many friends, conversationalist, social life, likes limelight, high social interest
- Low Score - few friends, shy, avoid social contacts, seclusiveness

E - EMOTIONAL STABILITY
- High Score - even mood, optimistic, cheerful, composed
- Low Score - moody, gloomy, pesimistic, daydreams, excitable, guilt and worry feelings

O - OBJECTIVITY
- High Score - thick skinned, less egoism
- Low Score - self-centered, suspicious, subjective, hypersensitive

F - FRIENDLINESS
- High Score - lack of fighting tendencies, pacifism, realistic way of treating frustrations, urge to please others, desire to be liked, tolerant of hostile action, accepts domination, respects others
- Low Score - hostility, fighting attitude, belligerent, resentful, wants to dominate, contempt for others

T - THOUGHTFULNESS
High Score - observing behavior of others, interest in thinking, philosophizing, mental poise, reflectiveness
Low Score - thoughtlessness, extraversion, likes overt acts, dislikes reflection

P - PERSONAL RELATIONS
High Score - tolerance and understanding of other people, faith in social institutions, good personal relations, cooperative
Low Score - fault finding, critical of other people and of institutions, hypercritical, suspicious, self-pity

M - MASCULINITY
High Score - not easily disgusted, not fearful, interest in masculine activities, hard-boiled, resistant to fear
Low Score - sympathetic, romatic, feminine activities, easily disgusted, fearful, emotional expressiveness

Just Suppose Inventory: Interpretation of scores

I. Attitude toward parents:

High - Parents do their best to understand children and do what is best for them; they appreciate the efforts of the school.
Low - Parents are too generous and too permissive with their children; they are unfairly critical of the school.

II. Attitude toward different size communities:

High - People are basically the same, regardless of the size of the community in which they live.
Low - Farm families are behind the times and crude. Small towns are dull. City people are unfriendly and critical of others.

III. Attitude toward divorce and families where mother works:

High - Acceptance of divorce as part of todays society. Recognition that parent-child relationships can be satisfactory if the mother works.
Low - Children from broken homes are usually delinquent. A mother's place is in the home. Divorced parents show little concern for their children.

IV. Attitude toward foreign born:

High - Generally, families with foreign born parents can make valuable contributions to our society.
Low - Foreigners tend to increase the crime rate and lower the standard of living.

V. Attitude toward persons with high or low educational backgrounds:

High - People are basically the same regardless of their educational level.
Low - Uneducated people are uninterested in the better things of life and do not cooperate with the school. Professional people are unwilling to accept those who work with their hands.

VI. Attitude toward low-income groups:

High - People in slum areas are victims of circumstance - they could do better if given a chance. It would be a challenge to try to help them.
Low - People in slum areas are lazy, indifferent and have low intelligence.

VII. Attitude toward different religions:

High - Religious beliefs are personal and differences should not influence ones acceptance of a person.
Low - It would be difficult to work with people whose religious beliefs differ from mine.

VIII. Attitude toward middle and upper-class groups:

High - It is the individual in the group that is important, not the class they're from.
Low - People from the upper-class lead an artificial life: lack interest in the school and "real family life". Middle-class families are too concerned about keeping up with the crowd; have little control over their children.

IX. Attitude toward teaching in a school with many low I. Q. and delinquent students and disinterested parents:

High - These students need encouragement and guidance; teaching would be a challenge.
Low - These students wouldn't behave or learn anything. Their parents are failures.

X. Attitude toward families of the laboring class:

High - Many are good-hearted, down to earth people who appreciate what the schools are doing.
Low - Laboring class families are dull. It would be undesirable to live near them.

XI. Attitude toward an ethnic group other than ones own:

High - The differences are not really important between my group and theirs.
Low - They have too many objectionable traits; parents do not care what their children do.

XII. Attitude toward a 3 generation family in a home:

High - Harmonious relationships can be achieved if family members respect each other.
Low - Old people are bossy and teen-agers are inconsiderate and noisy.

APPENDIX B. JUDGE'S RATING SHEET

JUDGE'S RATING SHEET

Case No. _____

_____
Evaluator

Directions:

Assume that this student will teach home economics in the following situation:

a relatively small Iowa town with a population of 1,200-7,000 in a community with few lower-class families
high school enrollment of 100-400 students
one teacher home economics department

Considering the evidence presented, what is your estimate of the chances that this student would become an effective high school teacher of home economics classes?

Indicate your opinion by writing a number from 0 - 10 in the space provided.

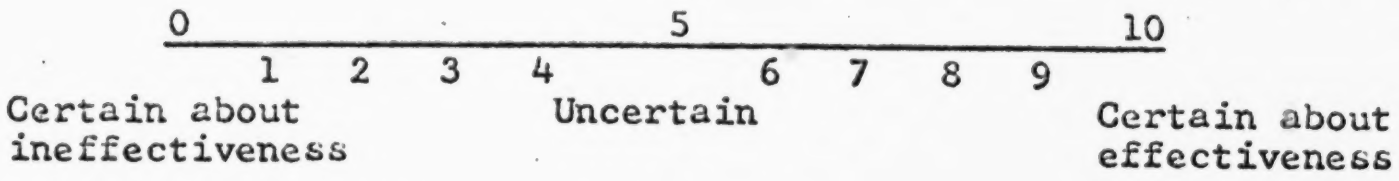
1) If you definitely think this student has the ability and personal qualities that she needs to become an effective teacher. Write 10 in the space provided.

2) If you definitely think this student does not have the personal qualities and/or ability to be an effective teacher, write 0 in the space provided.

3) Use numbers 1 to 9 to indicate another degree of certainty about her effectiveness or ineffectiveness. A response of 5 indicates you are uncertain, or the data are inadequate for a judgment.

JUDGE'S RATING SHEET (continued)

The following scale may help you keep these directions in mind.



Place your estimate in this box.

If your estimate is 5 or below, indicate your reason(s).

If this student were to teach in a large school system in an urban area, would you change your estimate of her success?
_____yes _____no

If yes, place your estimate in this box.

If this new estimate is 5 or below, indicate your reason(s).